21.1. Two-Way Classification

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk},\ i = 1, \ldots, I; j = 1, \ldots, J_i; k = 1, \ldots, K_{ij}$$

$$\varepsilon_{ijk} \sim N(0, \sigma^2),\ \text{independent.}$$

The least-squares estimates are $\hat{\mu}_{ij} = \bar{Y}_{ij\cdot} = K_{ij}^{-1} \sum_{k=1}^{K_{ij}} Y_{ijk}$.

Reparametrization in the balanced case:

balance: $J_i = J, K_{ij} = K$

$$\begin{aligned} \mu_{ij} &= \bar{\mu}_{\cdot\cdot} + (\bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot}) + (\bar{\mu}_{\cdot j} - \bar{\mu}_{\cdot\cdot}) + (\mu_{ij} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot j} + \bar{\mu}_{\cdot\cdot}) \\ &= \mu + \alpha_i + \beta_j + \gamma_{ij}, \end{aligned}$$

where

$$\begin{aligned} \mu &\equiv \bar{\mu}_{\cdot\cdot} = \text{overall mean}, \\ \alpha_i &\equiv \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot\cdot} = \text{effect of the } i\text{th level of factor } A, \\ \beta_j &\equiv \bar{\mu}_{\cdot j} - \bar{\mu}_{\cdot\cdot} = \text{effect of the } j\text{th level of factor } B, \\ \gamma_{ij} &\equiv \mu_{ij} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot j} + \bar{\mu}_{\cdot\cdot} = \text{effect of interaction between} \\ &\quad\ i\text{th level of factor } A \text{ and } j\text{th level of factor } B. \end{aligned}$$

Convenient and natural identifiability constraints are

$$\sum_{i=1}^{I} \alpha_i = 0,\ \sum_{j=1}^{J} \beta_j = 0,\ \sum_{i=1}^{I} \gamma_{ij} = \sum_{j=1}^{J} \gamma_{ij} = 0.$$

The least-squares estimates are

$$\hat{\mu} = \bar{Y}_{\cdots}, \quad \hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots}, \quad \hat{\beta}_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots},$$

$$\hat{\gamma}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots}.$$

21.2. Hypotheses of Interest

Hypothesis of no Interaction:

$$H_{AB} : \gamma_{ij} = 0, \ \forall \, i, j.$$

No interaction means the difference between two groups i, i' defined by factor A is independent of the level (j) of factor B. No interaction also means the difference between two groups j, j' defined by factor B is independent of the level (i) of factor A.

Hypotheses on Main Effects:

$$H_A : \alpha_i = 0, \ \forall \, i \ \ \text{(Main effect of } A\text{)}.$$

$$H_B : \beta_j = 0, \ \forall \, j \ \ \text{(Main effect of } B\text{)}.$$

Note: H_A is equivalent to

$$\bar{\mu}_{i\cdot} - \bar{\mu}_{i'\cdot} = 0, \ \forall \, i,$$

i.e., *averaged across levels of factor* B, the average mean is constant across levels of factor A.

If $H_{AB} : \gamma_{ij} = 0$ is true, this is equivalent to

$$\mu_{ij} - \mu_{i'j} = 0, \ \forall \, i, i', j,$$

i.e., the mean is constant across levels of factor A within each level of factor B. Thus, hypotheses on main effects make most sense if there is no interaction.

Test Statistics

ANOVA decomposition:

$$SS_{\text{TOT}} = SS_A + SS_B + SS_{AB} + SS_E,$$

where

$$\begin{aligned}
SS_{\text{TOT}} &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\cdots})^2, \\
SS_A &= \sum_i \sum_j \sum_k (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2 \\
&= \sum_i \sum_j \sum_k \hat{\alpha}_i^2 = RSS_{H_A} - RSS, \\
SS_B &= \sum_i \sum_j \sum_k (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})^2 \\
&= \sum_i \sum_j \sum_k \hat{\beta}_j^2 = RSS_{H_B} - RSS, \\
SS_{AB} &= \sum_i \sum_j \sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2 \\
&= \sum_i \sum_j \sum_k \hat{\gamma}_{ij}^2 = RSS_{H_{AB}} - RSS, \\
SS_E &= \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2 = RSS.
\end{aligned}$$

Proof:

$$\begin{aligned}\varepsilon_{ijk} &= \bar{\varepsilon}_{...} + (\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...}) + (\bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...}) \\ &\quad + (\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{...}) + (\varepsilon_{ijk} - \bar{\varepsilon}_{ij.})\end{aligned}$$

Cross-product terms are again 0,

$$\begin{aligned}\sum \varepsilon_{ijk}^2 &= \sum \bar{\varepsilon}_{...}^2 + \sum (\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2 + \sum (\bar{\varepsilon}_{.j.} - \bar{\varepsilon}_{...})^2 \\ &\quad + \sum (\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{...})^2 + \sum (\varepsilon_{ijk} - \bar{\varepsilon}_{ij.})^2\end{aligned}$$

where $\sum$ denotes $\sum_i \sum_j \sum_k$. Now substitute $\varepsilon_{ijk} = Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij}$, etc:

$$\begin{aligned}\sum (Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2 &= \sum (\bar{Y}_{...} - \mu)^2 + \sum (\bar{Y}_{i..} - \bar{Y}_{...} - \alpha_i)^2 \\ &\quad + \sum (\bar{Y}_{.j.} - \bar{Y}_{...} - \beta_j)^2 \\ &\quad + \sum (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} - \gamma_{ij})^2 \\ &\quad + \sum (Y_{ijk} - \bar{Y}_{ij.})^2.\end{aligned}$$

This justifies the least-squares estimates given above and $RSS = \sum \sum \sum (Y_{ijk} - \bar{Y}_{ij.})^2$.

If we consider minimizing RSS with, for example, $\gamma_{ij} = 0$, we get $RSS_{H_{AB}} =$

$$\begin{aligned}\sum (Y_{ijk} - \mu - \alpha_i - \beta_j)^2 &= \sum (\bar{Y}_{...} - \mu)^2 + \sum (\bar{Y}_{i..} - \bar{Y}_{...} - \alpha_i)^2 \\ &+ \sum (\bar{Y}_{.j.} - \bar{Y}_{...} - \beta_j)^2 \\ &+ \sum (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \\ &+ \sum (Y_{ijk} - \bar{Y}_{ij.})^2.\end{aligned}$$

We see that the other parameter estimates are the same. In fact, we can see that

$$\begin{aligned}&\sum (Y_{ijk} - \mu - \alpha_i - \beta_j)^2 = RSS_{H_{AB}} = \\ &= \sum (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + \sum (Y_{ijk} - \bar{Y}_{ij.})^2 \\ &= \sum (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 + RSS\end{aligned}$$

This shows that

$$RSS_{H_{AB}} - RSS = \sum_i \sum_j \sum_k (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2.$$

21.4. Distribution Theory

H_A, H_B, and H_{AB} are all of the form $\mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, where $\boldsymbol{\mu} = (\mu_{11}, \ldots, \mu_{IJ})'$, so the general theory of the F test applies.

What are the df?

For error: $df = n - p = IJK - IJ = IJ(K-1)$.

For a hypothesis $H : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, df=number of linearly independent restrictions on the means.

Degrees of freedom:

$$
\begin{aligned}
H_A\text{:} &\quad df = I - 1 \\
H_B\text{:} &\quad df = J - 1 \\
H_{AB}\text{:} &\quad df = (I-1)(J-1)
\end{aligned}
$$

μ_{11}	μ_{12}	$\cdots$	μ_{1J}	$\bar{\mu}_{1\cdot}$
μ_{21}	μ_{22}	$\cdots$	μ_{2J}	$\bar{\mu}_{2\cdot}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
μ_{I1}	μ_{I2}	$\cdots$	μ_{IJ}	$\bar{\mu}_{I\cdot}$
$\bar{\mu}_{\cdot 1}$	$\bar{\mu}_{\cdot 2}$	$\cdots$	$\bar{\mu}_{\cdot J}$	

$$
\begin{aligned}
F_A &= \frac{SS_A/(I-1)}{SS_E/[IJ(K-1)]} \sim F_{I-1, IJ(K-1)}, \text{ under } H_A \\
F_B &= \frac{SS_B/(J-1)}{SS_E/[IJ(K-1)]} \sim F_{J-1, IJ(K-1)}, \text{ under } H_B \\
F_{AB} &= \frac{SS_{AB}/[(I-1)(J-1)]}{SS_E/[IJ(K-1)]} \sim F_{(I-1)(J-1), IJ(K-1)}, \text{ under } H_{AB}
\end{aligned}
$$

ANOVA Table:

Source	df	SS	MS	F
A	$I-1$	$\sum_i \sum_j \sum_k (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots})^2$	$MS_A = \frac{SS_A}{I-1}$	$\frac{MS_A}{MS_E}$
B	$J-1$	$\sum_i \sum_j \sum_k (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots})^2$	$MS_B = \frac{SS_B}{J-1}$	$\frac{MS_B}{MS_E}$
AB	$(I-1)(J-1)$	$\sum_i \sum_j \sum_k (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2$	$MS_{AB} = \frac{SS_{AB}}{(I-1)(J-1)}$	$\frac{MS_{AB}}{MS_E}$
Error	$IJ(K-1)$	$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$MS_E = \frac{SS_E}{IJ(K-1)}$	
Total	$IJK-1$	$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{\cdots})^2$		

Unbalanced Two-Way ANOVA:

In the unbalanced case the orthogonal decomposition leading to the ANOVA decomposition does not hold. Parameter estimates of main effects now depend on whether interaction terms are in the model. Common prsctice is to first test the interaction and then test the main effects only if there is no significant interaction.